

# ParetoPrep: Fast computation of Path Skylines Queries

Michael Shekelyan   Gregor Jossé   Matthias Schubert  
 Institute for Informatics, Ludwig-Maximilians-Universität München  
 Oettingenstr. 67  
 D-80538 Munich, Germany  
 {shekelyan,josse,schubert}@dbs.ifi.lmu.de

## ABSTRACT

Computing cost optimal paths in network data is a very important task in many application areas like transportation networks, computer networks or social graphs. In many cases, the cost of an edge can be described by various cost criteria. For example, in a road network possible cost criteria are distance, time, ascent, energy consumption or toll fees. In such a multicriteria network, a route or path skyline query computes the set of all paths having pareto optimal costs, i.e. each result path is optimal for different user preferences. In this paper, we propose a new method for computing route skylines which significantly decreases processing time and memory consumption. Furthermore, our method does not rely on any precomputation or indexing method and thus, it is suitable for dynamically changing edge costs. Our experiments demonstrate that our method outperforms state of the art approaches and allows highly efficient path skyline computation without any preprocessing.

## 1. INTRODUCTION

In recent years, querying network data has become more and more important in many application areas like transportation systems, the world wide web, computer networks or social graphs. One of the most important tasks in network data is computing cost optimal paths between two nodes. Especially in transportation networks and computer networks finding shortest or cost optimal paths is essential for optimizing the movement of objects or information. An optimal path can depend on multiple cost criteria and a network considering different cost dimensions for each link is called a multicost or multicriteria network. In road networks, possible optimization criteria are travel time, travel distance, toll fees, environmental hazards or energy consumption. In computer networks, typical cost criteria are bandwidth, rental cost and current traffic.

A simple solution to compute optimal paths in multicriteria networks is to combine all cost values into a single opti-

mization criterion. However, finding a suitable combination method is a difficult task because a good choice might depend on personal preferences and the current query context.

An alternative approach is to compute all pareto optimal paths, i.e. all paths that are potentially cost optimal under any cost combination. In the database community, this task is generally known as skyline query [1] and when looking for routes in a network as route skyline query [11]. In other communities like Artificial Intelligence and Operations Research the latter task is known as multicriteria shortest path optimization [21].

One of the most efficient approaches to answering route skyline queries is the label correcting approach which exploits the pareto optimality of any subpath of a result path. However, for large graph this pruning mechanism is not enough to ensure efficient computation because the majority of the network has to be visited. To efficiently compute path skylines in large networks, it is necessary to direct the search towards the target, e.g. by using lower bound approximations for the remaining cost in each criteria in a similar way as A\*-search does in the single criterion case.

For example, a lower bound approximation for network distance is the Euclidean distance. Considering multiple and arbitrary cost criteria requires a more general approach for acquiring lower bound approximations. One solution is precalculating bounds like a Reference Node Embedding [6]. However, precalculated bounds have several drawbacks. Firstly, the approximation quality is often insufficient for the majority of queries. Secondly, they typically require a lot of memory compared to the size of the network. And finally, in dynamic networks – where edge costs vary over time – the approximation may lose its bounding properties.

Alternatively, [27] makes use of the following skyline property: For every particular cost criterion there exists an optimal path and each nondominated path cannot have a smaller cost value for this criterion. Now, by performing a single-source all-target Dijkstra search, it is possible to compute the single criterion optimal paths from each node in the network to the target node and thus, derive lower bounds for each cost value w.r.t. the considered cost criterion. Performing this search for all cost criteria, computes a vector of lower bounds for each node. Although the effort is large, the quality of the bounds compensates for the overhead when computing the path skyline later on. However, the runtime degenerates for large graphs and multiple criteria because the entire graph is visited once for each criterion.

To avoid the drawbacks of both approaches, we propose a new method, *ParetoPrep*, which strongly increases the ef-

efficiency even for multiple cost criteria. Our method does not rely on preprocessing, hence it is applicable to networks with dynamically changing costs. ParetoPrep computes tight lower bound approximations which yield great pruning power and low runtime. Additionally, ParetoPrep finds the shortest paths between a start and a target node with respect to all  $d$  cost criteria within a single graph traversal. When traversing the graph ParetoPrep visits only a small part of the network, keeping the search local. To even further reduce the explored part of the graph, we introduce a bidirectional optimization of ParetoPrep.

In our experimental evaluation, we show that our new approach outperforms the state of the art in computing route skylines, i.e. the above mentioned approach of computing lower bounds through multiple Dijkstra searches [27] as well as the ARSC [11]. As test network, we use the open street map road network of the German federal state of Bavaria and examine searches for up to five different cost criteria. To conclude, the contributions of this paper are:

- We introduce an algorithm, ParetoPrep, which computes tighter lower bound cost approximations than any state of the art approach.
- We show that ParetoPrep outperforms any comparable algorithm in terms of runtime, memory space and effectiveness.
- We present a complete algorithm for answering path skyline queries on large and possibly dynamic network data and a bidirectional optimization of this algorithm.

The rest of the paper is organized as follows: In Section 2 we present related work in the area of path skyline computation and routing in large networks. Section 3 provides basic notations, concepts and formalizes path skyline queries. ParetoPrep, our new algorithm, is presented in Section 4. Additionally, the section contains formal proofs concerning the correctness of ParetoPrep. Section 5 describes the results of our experiments, demonstrating the various benefits of ParetoPrep. The paper concludes with a summary and directions for future work in Section 6.

## 2. RELATED WORK

The task of finding all pareto optimal elements within a vector space was introduced to the database community as skyline operator in [1]. Since then, multiple methods for computing skylines in database systems on sets of (cost) vectors followed [24, 10, 18].

With regard to network data and especially spatial networks there exist some approaches for computing skylines on landmarks or other points of interest. In [3] the authors introduce a method for calculating a skyline of landmarks in a road network which are compared w.r.t. their network distance to several query objects. In [8] the authors propose in-route skyline processing in road networks. Assuming that a user is moving along a predefined route to a known destination, the algorithm processes minimal detours to sets of landmarks being distributed along the path. In [9] the authors discuss continuous skyline queries in road networks, i.e. a moving user queries for a skyline of points of interest.

The task of computing a skyline of pareto optimal paths between two nodes is referred to as route skyline query in [11]. However, in accordance the graph theoretical terminology, we refer to it as path skyline query. In the following, we

will discuss solutions from other areas of research and discuss the differences between the algorithm ARSC proposed in [11] and alternative approaches.

In Operations Research, the problem of finding complete path skylines is known as the Multiobjective Shortest Path problem and surveys on existing solutions to this problem can be found in [21, 19, 25, 5]. Early on, [7] proved that the size of the path skyline may increase exponentially with the number of hops between start and target node, and that the problem therefore is NP-hard. More recently, [16] showed that the number of routes is in practice feasibly low when using strongly correlated cost criteria.

The potential complexity of the problem motivated solutions which approximate the pareto front. A fully polynomial-time approximation scheme is provided by [7], [28] and [26]. Many approximative solutions are also based on genetic algorithms [17].

Concerning exact solutions, the most common approach are labeling algorithms which label nodes with the cost vectors of assembled paths ending at that node. They start with paths consisting of the outgoing edges of the start node, and in each iteration extend all previously assembled paths that may be part of a skyline path and terminate once all assembled paths were either extended or pruned. Paths can be pruned if they are dominated by a path with the same terminal node because all subpaths of nondominated paths are also nondominated. We will refer to this type of pruning as *local domination*. To perform local domination checks, a route skyline is maintained for each node, which consists of all so far nondominated routes ending at that node.

Labeling approaches can be subdivided into label setting and label correcting approaches. In each iteration, label setting approaches like Martin’s algorithm [15] extend a non-dominated path and therefore have a well-defined worst case performance, limited by the number of nondominated paths. In contrast, label correcting approaches do not necessarily extend nondominated paths which offers more flexibility. For instance, the label correcting approach allows to extend the complete route skyline of a node in each iteration, instead of separately extending particular paths.

In the special case of two cost criteria, node skylines can be sorted to improve the efficiency of local domination checks. Two dimensional skyline cost vectors sorted in ascending order by the first criterion are simultaneously sorted in descending order w.r.t. to the second criterion. When the route skyline of node  $u$  is extended with edge  $(u, v)$ , it has to be merged with the set of nondominated paths ending at  $v$ . [2] shows how to efficiently solve the problem of merging two sorted route skylines and [22] describes an efficient method to check if one skyline completely dominates the other making the merging step trivial.

The labeling algorithms cited so far solve the single-source problem of finding all nondominated paths from  $s$  to all other nodes. In [23], it is shown that when solving the point-to-point problem from a start node  $s$  to a target node  $t$  all partially or fully extended paths dominated by a path from  $s$  to  $t$  can be pruned. We will refer to this type of pruning as global domination. Global domination allows to process only a part of the graph and significantly reduces the number of considered paths to compute the path skyline. Thus, to answer path skyline queries in large graphs global domination should be employed

to direct the search. [23] also showed that lower bound

cost estimations of fully extended paths can be utilized to considerably improve the pruning power of global domination and can be used to prioritize the extension of more promising paths.

For cost criteria other than distance, lower bound cost estimations are usually acquired through some kind of pre-processing. In ARSC [11], lower bounds are provided by a Reference Node Embedding computed before any query is posed. However, a reference point embedding usually does not yield good lower bounds for all possible queries. Furthermore, the memory consumption is rather large and any precomputed information has to be checked for validity in case of dynamically changing edge costs. Another approach is to compute lower bound costs individually for each query [13]. Such lower bound cost computation may be combined with computing the shortest paths for each criterion which can then be used for global domination checks. Tung and Chew [27] proposed to perform a single-source all-target Dijkstra search for each cost criterion on the graph in reverse direction [4] to find the costs of the shortest paths from all other nodes in the graph to the target node  $t$ . We will refer to this approach as Multidijkstra (MD). A major shortcoming of this approach is that it processes the whole graph for every cost criterion.

In the special case of two cost criteria, Machuca and Mandow [14] propose a termination condition which makes use of special properties of two dimensional skylines. We would like to stress that this method is not applicable to the case of more than two cost criteria. Since the algorithm in [14] uses two Dijkstra searches, we shall refer to it as Doubledijkstra (DD). In [29] the authors compute a set of optimal paths in time-dependent, uncertain, multicriteria networks. The presented model allows the complex representation of a dynamic network model. However, the lower bound computation relies on the same undirected search as in [27].

It should be stressed that MD and DD do not perform point-to-point shortest path searches. Therefore, speed-up techniques for point-to-point searches are not applicable.

The methods presented in this paper use significantly less runtime and memory space compared to the MD and DD approaches by restricting the visited part of the network and deriving the required lower bounds in a single graph traversal. Thus, Pareto Prep can be combined with any search algorithm requiring time and memory efficient computation of lower bound costs for multiple attributes.

### 3. PRELIMINARIES

#### 3.1 Route Skyline Queries

A multicost network is represented by a directed weighted graph  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$  comprised of a set of nodes  $\mathcal{V}$  and a set of directed edges  $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ . Each edge  $(n, m) \in \mathcal{E}$  is labeled with a cost vector  $cost(n, m) \in \mathcal{C} \subset \mathbb{R}_+^d$  which consists of the costs for traversing edge  $(n, m)$  w.r.t. each of the  $d$  cost criteria (in mathematical context also referred to as cost dimensions). If there exists an edge  $(n, m)$  then  $n$  and  $m$  are neighboring nodes and  $(n, m)$  shall be called an outgoing edge of  $n$  and an incoming edge of  $m$ .

A sequence of adjacent edges connecting two nodes  $s$  and  $t$ ,  $w = ((s, n_1), (n_1, n_2), \dots, (n_k, t))$ , is called a way from  $s$  to  $t$ . If  $w$  does not visit any node twice it is called a path or a route. The cost of a path  $p$  in the  $i$ -th cost dimension is

the sum of its individual edge costs:

$$cost(p)_i = \sum_{(n, m) \in p} cost(n, m)_i$$

A cost vector  $a \in \mathbb{R}_+^d$  dominates a cost vector  $b \in \mathbb{R}_+^d$ , denoted  $a \prec_{\text{dom}} b$ , iff  $a$  has a smaller cost value than  $b$  in at least one dimension  $i$  and  $b$  does not have a smaller cost value than  $a$  in any dimension  $j$ :

$$\exists_{i \in \{1, \dots, d\}} : a_i < b_i \wedge \nexists_{j \in \{1, \dots, d\}} : a_j > b_j \Leftrightarrow a \prec_{\text{dom}} b$$

For a set of paths  $A$ ,  $A \prec_{\text{dom}} b$  implies that there exists some path  $a \in A$  which dominates the path  $b$ . Cost vectors which are not dominated by any other cost vector are called nondominated or pareto optimal. The set of nondominated cost vectors includes all optimal trade-offs between cost criteria and is called the pareto front or skyline [1].

**DEFINITION 1 (PATH SKYLINE).** Let  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$  be a multicost network, let  $s \in \mathcal{V}$  be the start node and let  $t \in \mathcal{V}$  be the target node. Then, the set of all paths between  $s$  and  $t$  whose cost vectors are nondominated is defined as path skyline  $\mathcal{S}(s, t)$ .

The task we examine in this paper is to efficiently compute  $\mathcal{S}(s, t)$ . In the following, we always assume a multicost network is given.

#### 3.2 Label Correction and Pruning

Our solution employs the label correcting approach and incorporates two domination checks in order to prune paths from the search tree. The first type of domination is local domination which is used to compare paths starting and ending at the same node. The second domination relationship is global domination it describes whether a path still can be extended to a nondominated path between the start and the target node of the given query.

**DEFINITION 2 (LOCAL DOMINATION).** Let  $p$  and  $q$  be two paths starting at the same node  $s$  and ending at the same node  $t$ . If  $cost(p) \prec_{\text{dom}} cost(q)$  holds, we refer to  $q$  as dominated by  $p$  :  $p \prec_{\text{dom}} q$ . Correspondingly,  $q$  is referred to as nondominated iff  $\nexists p = ((s, n_1), \dots, (n_k, t)) : cost(p) \prec_{\text{dom}} cost(q)$ .

For any path  $q = ((n_1, n_2), \dots, (n_{l-1}, n_l))$ , any subsequence of edges  $p = ((n_i, n_{i+1}), \dots, (n_{k-1}, n_k))$  is called a subpath of  $q$ . The following lemma shows that locally dominated paths can be pruned from the search.

**LEMMA 1 (LOCAL DOMINATION CHECK).** Any subpath of a nondominated path is nondominated (w.r.t. its start and target node).

**PROOF.** Suppose, a path  $p$  had a dominated subpath  $q$ . This subpath could be substituted with the dominating path  $q'$ . The modified path  $p'$  would then dominate the original path  $p$ :

$$cost(p') = cost(p) - cost(q) + cost(q')$$

$$cost(q') \prec_{\text{dom}} cost(q) \Rightarrow cost(p') \prec_{\text{dom}} cost(p)$$

This contradicts the assumption and thereby proves the claim.

□

Using local domination allows to decide which paths between node  $s$  and some other node  $n$  have to be extended when processing node  $n$ . However, it is sufficient to decide whether any path between  $s$  and  $n$  can be extended into a nondominated path between  $s$  and  $t$ . To test this characteristic, we introduce global domination.

**DEFINITION 3 (GLOBAL DOMINATION).** *Given a start node  $s$  and a destination node  $t$ , an arbitrary path  $p$  is called globally dominated iff it is a subpath of a dominated path  $q$  between  $s$  and  $t$ . Respectively, any subpath  $p$  of a nondominated path  $q$  between  $s$  and  $t$  is called globally nondominated.*

Due to global domination it is not necessary to extend any path  $p$  between  $s$  and some other node  $n$ , if there does not exist any extension  $q$  of  $p$  ending at the target  $t$  and being part of  $\mathcal{S}(s, t)$ . Without the

concept of global domination an algorithm cannot stop the search until all path skylines  $\mathcal{S}(s, v)$  for all  $v \in \mathcal{V}$  are calculated. To exploit the fact that we are only interested in  $\mathcal{S}(s, t)$ , a simple global domination check is to compare the cost of any path  $p$  to  $\mathcal{S}(s, t)$ . If  $\exists q \in \mathcal{S}(s, t) : \text{cost}(q) \prec_{\text{dom}} \text{cost}(p)$  then  $p$  cannot be extended into an element of  $\mathcal{S}(s, t)$  because  $\text{cost}(p)$  cannot be decreased by extending  $p$  for any cost criterion  $i$ . This basic global domination check is equivalent to the stopping criterion of Dijkstra's algorithm for single criterion shortest path computation. However, the check is not sufficient to restrict the search space for the path skyline to a limited portion of the network. To check for global domination in a more effective way, lower bound cost approximations are employed. Before explaining the global domination check being employed in this paper, we need to introduce some notation:

$\max(a, b)$  shall denote the component-wise (criterion-wise) maximum of cost vectors  $a$  and  $b$ , i.e.  $(\max(a, b))_{1 \leq i \leq d} := \max\{a_i, b_i\}$ .

**DEFINITION 4 (LOWER BOUND COSTS).** *Let  $n$  and  $m$  be nodes and let  $i$  be an arbitrary cost criterion. If for the real value  $lb(a, b)_i$  holds  $lb(a, b)_i \leq \text{cost}(p)_i$  for any path  $p$  connecting  $a$  and  $b$  then  $lb(a, b)_i$  is called lower bound for  $i$  w.r.t.  $a$  and  $b$ . The vector consisting of the lower bounds of all cost criteria w.r.t.  $a$  and  $b$  is denoted as  $lb(a, b)$ .*

To compute lower bounds, it is possible to employ the MultiDijkstra (MD) method which is computing all single criterion optimal paths starting at some node  $n$  and ending at node  $t$  as described in section 2. In the next section, we will introduce our new method ParetoPrep which is computing even tighter lower bounds in much more efficient way. Additionally, to computing lower bound cost both algorithms compute the optimal paths between the start node  $s$  and the target node  $t$  for each cost criterion  $i$ . Both information are employed to prune paths from the search space in the following domination check:

**LEMMA 2 (GLOBAL DOMINATION CHECK).** *Let  $p$  be a path from  $n$  to  $m$  and  $q$  be a path from the start node  $s$  to the target node  $t$ . If*

$$\text{cost}(q) \prec_{\text{dom}} \max(lb(s, n) + \text{cost}(p) + lb(m, t), lb(s, t))$$

*, then the path  $p$  is globally dominated.*

**PROOF.** The cost vector  $lb(s, n) + \text{cost}(p) + lb(m, t)$  is the lower bound cost of all paths from  $s$  to  $t$  via  $p$ . If this lower

bound cost is dominated by the cost of a path from  $s$  to  $t$ , there does not exist a nondominated path from  $s$  to  $t$  via  $p$ . Additionally, there cannot be any path  $q$  from  $s$  to  $t$  where  $\text{cost}(q)_i$  is smaller than the cost of the shortest path w.r.t. to criterion  $i$ . Thus, the bound can always be increased to the cost of the shortest path in dimension  $i$ .  $\square$

Now that we have introduced our pruning criteria, let us explain the label correcting algorithms being used in the paper: Any label correcting approach employs two important data structures. The first is a node table maintaining an entry for each visited node. The entry of node  $n$  maintains a local skyline, i.e. a list of nondominated paths starting in  $s$  and ending in  $n$ . For each path a flag is maintained, denoting whether the path has already been processed or not. The second data structure is a queue of node entries which is ordered with respect to a preference criterion. In our implementation, we use the sum over the minimal cost values of the local skyline in each cost criterion, but other preference function work as well.

When answering a path skyline query for a given start node  $s$  and a target node  $t$ , the search starts by visiting the outgoing edges of  $s$ . For each newly reached node, a node tab entry is generated and added to the queue. In the main loop of the algorithm the top entry is removed from the queue. Any path in its local skyline which has neither been processed nor is globally dominated by any result path found so far is extended. That means, the path is extended by all outgoing links of node  $n$ . A new path is registered in the local skyline of the respective target node if it is not dominated by the other paths in that entry. If a node tab entry is updated with a new path it is added to the queue. The algorithm terminates when the queue is empty, i.e. no entry contains any unprocessed path. This algorithm is similar to the ARSC algorithm introduced in [11]. However, the global domination check in ARSC is less strong because it cannot exploit the existence of the single criterion shortest paths. Furthermore, the lower bounds are less tight due to the use of the precomputed Reference Node Embedding.

### 3.3 Special Case: Two Cost Criteria

Although we focus on the general multicriteria case, we want to point out there exist two major optimizations for the special case of two cost criteria. One pertains to the problem of skyline merging, the other significantly limits the number of nodes that have to be visited. Both optimizations stem from the same quality of two-dimensional skylines: Sorting cost vectors w.r.t. one criterion implies a sorting in reversed order w.r.t. the other criterion.

This can be used to more efficiently maintain local skylines in sorted lists [2]. As a result, the only paths potentially dominating a path  $p$  are

its two neighbors in the sorted skyline. Furthermore, the set of potentially dominated skyline paths is also restricted to neighboring paths [20]. We employ this optimization for all compared algorithms in the experiments using only two cost criteria.

The above also implies that the optimal path w.r.t. one criterion (by definition part of the skyline) is the weakest skyline path w.r.t. to the other criterion. Hence, if  $s^1$  is the optimal path from  $s$  to  $t$  w.r.t. the first criterion, then  $\text{cost}(s^1)_2$  constitutes the upper bound for the second criterion, and vice versa.

```

// step 1: initialisation
foreach outgoing edge (s, m) of s do
    S(m) ← S(m) ∪ path consisting of the edge (s, m)
    open ← open ∪ {m}

// step 2-4: main loop
while open ≠ {} do
    // step 2: node selection
    n ← arg mino ∈ open (minp ∈ S(o) ∑i=1d cost(p)i + lb(o, t)i)
    open ← open \ {n}

    // step 3: mark globally dominated as processed
    foreach unprocessed p ∈ S(n) do
        if S(t) <dom max(lb(s, t), cost(p) + lb(n, t)) then
            mark p as processed

    // step 4: extend unprocessed paths in S(n)
    A ← unprocessed paths in S(n)
    foreach outgoing edge (n, m) of n do
        B ← paths in A extended with (n, m)
        S(m) ← merged skyline of S(m) and B
        if S(m) was modified then
            open ← open ∪ {m}
    mark all paths in S(n) as processed

// step 5: termination
return S(t)

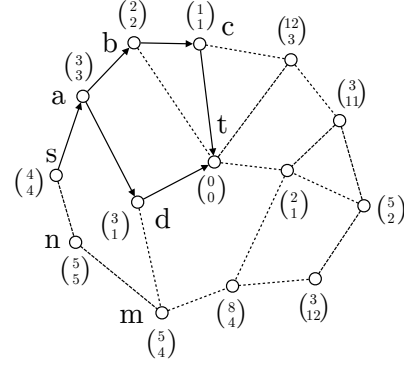
```

**Algorithm 1:** Pseudocode of ARSC, a label-correcting search for all nondominated paths from  $s$  to  $t$ .

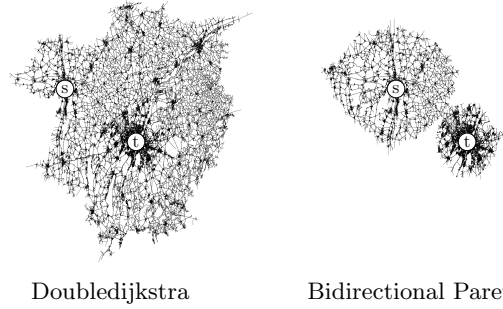
This property can be used to restrict the search space of lower bound computation of the MD approach for the two criteria case and we will refer to this approach as Doubledijkstra (DD) [14]. The optimization starts with a reverse dijkstra search for the shortest path from  $t$  to  $s$  w.r.t. the first criterion. After this search reaches  $s$ , the search is done w.r.t. the second criterion. Thus, we have upper bounds for both criteria. Finally, the search for both criteria is continued until all nodes having one cost value being smaller than the corresponding upper bound are found. Let us note that this optimization is only applicable to the MD approach and it is still significantly less efficient than our new method ParetoPrep due to lower quality of the lower bounds.

## 4. LOWER BOUND COMPUTATION

In this section, we introduce our novel approach ParetoPrep which computes all required lower bounds for efficiently computing the path skyline  $S(s, t)$ . The idea of ParetoPrep is to compute all single criterion shortest paths between  $s$  and  $t$  within a single partial graph traversal. It will be shown that there cannot be a nondominated path between  $s$  and  $t$  containing any edge that is not visited during this traversal. Thus, ParetoPrep visits all required nodes and edges for processing a path skyline query. At the end of the section, we will introduce bidirectional ParetoPrep which further limits the nodes being visited during the lower bound computation.



**Figure 1:** Exemplary output of ParetoPrep given a start node  $s$  and a target node  $t$ . The indicated paths  $\{s, a, b, c, t\}$  and  $\{s, a, d, t\}$  are the shortest paths for the first and second criterion. The vectors next to each node are the computed lower bound costs  $lb$  of reaching  $t$  from the respective nodes.



**Figure 2:** Comparison of visited nodes by Doubledijkstra and Bidirectional ParetoPrep for a routing task from Augsburg to Munich with  $dur/ener$  as cost criteria.

### 4.1 ParetoPrep

The goal of a precomputation step such as ParetoPrep or MD is to compute a lower bound for each cost criterion and each node that is potentially visited when computing the path skyline with a label correcting algorithm. The cost of the shortest path w.r.t. each cost criterion represents an optimal bound. The shortest paths are always part of the route skyline and any additional path being part of the skyline cannot have any cost value being smaller than the shortest path w.r.t. the corresponding cost criterion. The MD approach performs separate searches for each cost criterion and visits the entire network for each criterion.

The idea of ParetoPrep is to avoid traversing the same subpaths multiple times by computing the bounds in a single traversal of the relevant portion of the graph. Furthermore, ParetoPrep limits the search space due to the usage of a global domination check.

In this subsection, we will describe ParetoPrep and the information it computes. In the next subsection, we will show why this resulting information allows a fast computation of the path skyline. We will start our discussion by giving an overview of the used data structures. ParetoPrep maintains a set of open nodes  $open$  and a set  $S$  of paths from  $s$  to  $t$ .

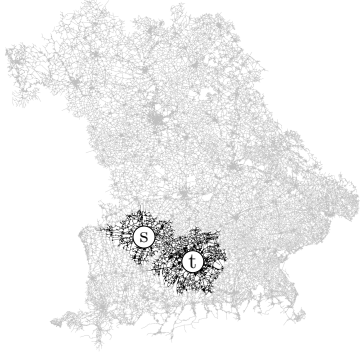


Figure 3: Comparison of visited nodes by Bidirectional ParetoPrep (black subgraph) and Multidijkstra (complete graph in gray) for a routing task from Augsburg (s) to Munich (t) with all 5 criteria.

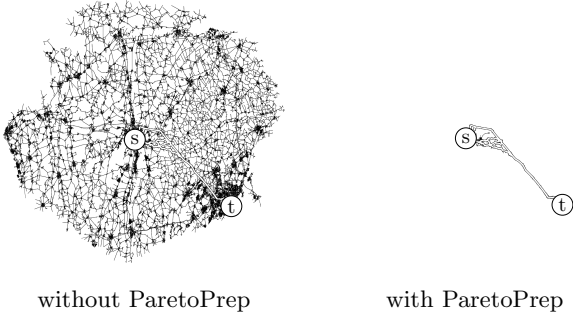


Figure 4: The comparison of label correcting search areas for a routing task from Augsburg (s) to Munich (t) with two cost criteria shows that the label correcting search considers almost no dominated paths when using the information provided by ParetoPrep.

Each visited node  $n$  has an entry consisting of two vectors  $\{lb(n, t), succ(n)\}$ .

The cost vectors  $lb(n, t) \in \mathbb{R}_+^d$  are the lower bound costs of all paths from  $n$  to  $t$ , through which  $n$  was previously reached in ParetoPrep. Upon termination of the algorithm  $n$  has been reached by all globally nondominated paths from  $n$  to  $t$ . The edges  $\succ n \in \mathcal{E}^d$  are the first edges of the currently shortest path from  $n$  to  $t$  w.r.t. criterion  $i$ . These successor edges are used to reconstruct the shortest paths from  $s$  to  $t$  w.r.t. to each of the cost criteria.

An entry of an unvisited node  $n$  is assumed to be  $lb(n, t)_i = \infty$ ,  $succ_i(n) = \emptyset$ .  $lb(t, t)_i$  is the zero vector because the lower bound costs of reaching  $t$  from  $t$  are zero.

The pseudocode of the algorithm is provided in Algorithm 2. Let  $s$  be the start node of the routing task,  $t$  the target node and  $d$  the number of cost criteria.

The first step of the algorithm is the initialization. The open set is created and the target node  $t$  is added to the open set. The second step is node selection. In each iteration, an open node  $n$  is selected and removed from the open set. To reduce the number of nodes which have to be visited twice, the nodes closest to  $t$  should be visited first. To achieve this, each node is ranked by the linear sum over the cost vector

```

// step 1:  initialisation
 $\mathcal{S} \leftarrow \{\}$ 
 $open \leftarrow \{t\}$ 
// step 2-5:  main loop
while  $open \neq \{\}$  do
    // step 2:  node selection
     $n \leftarrow \arg \min_{q \in open} \sum_{i=1}^d lb(q, t)_i$ 
     $open \leftarrow open \setminus \{n\}$ 
    // step 3:  global domination
    if  $\mathcal{S} \prec_{dom} lb(n, t)$  then
        skip step 4 and 5
    // step 4:  node expansion
    foreach incoming edge  $(m, n)$  of  $n$  do
        foreach criterion  $i$  do
            if  $lb(n, t)_i + cost(m, n)_i < lb(m, t)_i$  then
                 $lb(m, t)_i \leftarrow lb(n, t)_i + cost(m, n)_i$ 
                 $succ_i(m) \leftarrow (m, n)$ 
                 $open \leftarrow open \cup \{m\}$ 
    // step 5:  path construction
    if  $lb(s, t)$  was modified in step 4 then
        foreach modified component  $i$  of  $lb(s, t)$  do
             $p \leftarrow reconstructpath(s, t, succ_i, i)$ 
             $\mathcal{S} \leftarrow \mathcal{S} \cup \{p\}$ 
            remove paths dominated by  $p$  from  $\mathcal{S}$ 
RETURN  $\mathcal{S}$  AND  $lb$ 

```

Algorithm 2: Pseudocode of ParetoPrep

and the node with the smallest value is selected first.

The third step is a check whether the selected node  $n$  has to be extended. In other words, we compare the current lower bound  $lb(n, t)$  to the current set of shortest paths  $\mathcal{S}$  from  $s$  to  $t$ . If the vector  $lb(n, t)$  is globally dominated, the node does not need to be extended. This means, there cannot be any path from  $s$  to  $t$  via  $n$  which is not dominated by the already found shortest paths in  $\mathcal{S}$ .

If  $lb(n, t)$  is not globally dominated, we perform step four and five. The fourth step is the extension of the selected node  $n$ . In this step, we consider the neighboring nodes having a directed edge ending at  $n$ , i.e. the predecessors of  $n$ . The cost of each predecessor node  $m$  for each cost dimension  $i$  is set to the minimum of  $lb(m, t)_i$  and  $lb(n, t)_i + cost(m, n)_i$ . For each criterion  $i$  in which  $c(m)_i$  is changed, the  $i$ -th predecessor edge  $succ_i(m)$  is set to  $(m, n)$ . If  $lb(m, t)$  was changed and  $m$  is not the start node  $s$ ,  $m$  is added to the set of open nodes.

The fifth step is the construction of paths from  $s$  to  $t$ . Let us note that  $\mathcal{S}$  contains the set of shortest paths w.r.t. to each cost criterion and thus,  $\mathcal{S}$  is a subset of the route skyline being computed in the following label correcting search. This path construction step is only performed if at least one component of the vector  $lb(s, t)$  was modified in the previous step. For each modified cost criterion the currently short-

**Data:**  $s, t, succ_i, i$   
**Result:** Current shortest path from  $s$  to  $t$  for criterion  $i$   
 $m \leftarrow s$   
 $p \leftarrow$  new empty path  
**while**  $m \neq t$  **do**  
     $(m, n) \leftarrow succ_i(m)$   
     $p \leftarrow p$  extended with  $(m, n)$   
     $m \leftarrow n$   
**return**  $p$

**Algorithm 3:** Pseudocode of ParetoPrep's path construction routine

est path from  $s$  to  $t$  is constructed. These paths are constructed by following the respective successors  $succ_i$ , similarly to how paths are reconstructed in Dijkstra's search. The pseudocode is provided in Algorithm 3.

If after an iteration there are no more open nodes, the algorithm terminates. Upon termination,  $\mathcal{S}$  contains a shortest path from  $s$  to  $t$  for each cost criterion. And for every node  $n$  which could be part of a skyline path from  $s$  to  $t$ ,  $lb(n, t)$  contains the lower bound costs of reaching  $t$ .

Figure 4.1 illustrates an exemplary execution of the algorithm for a simple search task.

## 4.2 Formal Aspects of ParetoPrep

To show that ParetoPrep is a correct preprocessing step for computing path skylines, we will show that ParetoPrep computes valid lower bounds for all nodes that have to be visited during the search. Furthermore, we show that all single criterion shortest paths between  $s$  and  $t$  are found. We will start by introducing the relevant subgraph of  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$  which is specific for a path skyline query between  $s$  and  $t$ .

**DEFINITION 5 (SOLUTION GRAPH).** Let  $\mathcal{E}_{SOL}(s, t) \subseteq \mathcal{E}$  denote the set of all edges contained in some  $p \in \mathcal{S}(s, t)$ . Furthermore, let  $\mathcal{V}_{SOL} := \{v \in \mathcal{V} \mid v \in \mathcal{E}_{SOL}(s, t)\}$  and  $\mathcal{C}_{SOL} := \mathcal{C}|_{\mathcal{E}_{SOL}(s, t)}$ . The Solution Graph is defined as  $G_{SOL}(s, t) := (\mathcal{V}_{SOL}, \mathcal{E}_{SOL}(s, t), \mathcal{C}_{SOL})$ .

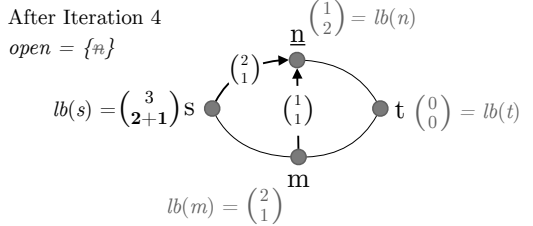
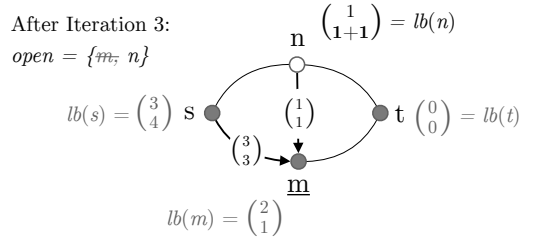
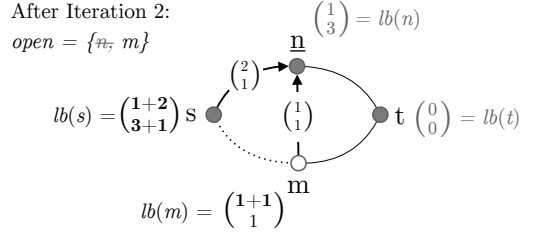
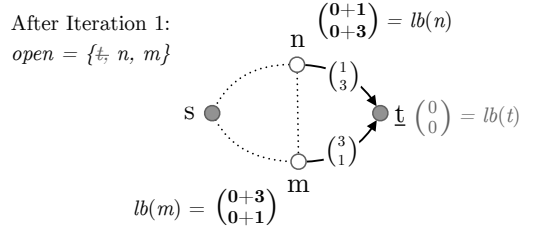
There are two important properties of  $G_{SOL}(s, t)$  making it the central concept for examining the correctness of ParetoPrep:

- It is not necessary to visit any edge  $e \notin \mathcal{E}_{SOL}(s, t)$  during the search.
- The cost vector of single criterion shortest paths in  $G_{SOL}(s, t)$  represent a viable lower bound for computing the path skyline.

The first property follows from the definition of  $G_{SOL}(s, t)$ . The second property is shown by the following lemma:

**LEMMA 3.** Let  $p = ((s, n_1), \dots, (m, n), \dots, (n_l, t)) \in \mathcal{S}(s, t)$  be a nondominated path, let  $q = ((s, n_1), \dots, (m, n))$  be a sub-path of  $p$  and let  $lb(n, t)_{SOL}$  be the vector of the single criterion shortest paths between  $n$  and  $t$  for each criterion in  $G_{SOL}(s, t)$ . Then  $cost(q) + lb(n, t)_{SOL}$  cannot be dominated in  $\mathcal{S}(s, t)$ .

**PROOF.** The single criterion shortest paths between  $n$  and  $t$  in  $G_{SOL}(s, t)$  lower bounds any path between  $n$  and  $t$  in  $G_{SOL}(s, t)$ .  $p$  is nondominated and for all  $1 \leq i \leq d$ :  $cost(p)_i \leq lb(n, t)_{SOL} + cost(q)_i$ . Thus,  $cost(q) + lb(n, t)_{SOL}$  cannot be dominated in  $\mathcal{S}(s, t)$ .  $\square$



**Figure 5: Exemplary execution of ParetoPrep.** The selected node of the iteration is underlined. After iteration 2 the path through  $[s, n_1, t]$  with the costs  $[3, 4]$  and after iteration 3 the path through  $[s, n_1, n_2, t]$  with the costs  $[6, 3]$  is constructed. ParetoPrep terminates after iteration 4.

We can now show two properties of ParetoPrep which prove the correctness of our approach:

- ParetoPrep visits every edge in  $G_{SOL}(s, t)$ .
- For every node  $n \in G_{SOL}(s, t)$ :  $lb(n, t)_i \leq cost(p_i)_i$ , where  $p_i$  is the shortest path in  $G_{SOL}(s, t)$  w.r.t. cost dimension  $i$ .

The first property implies that every necessary node or edge is visited. The second property implies that the lower bounds for the nodes relevant to the solution are correct. Note that for any node  $n \notin G_{SOL}(s, t)$   $lb(n, t)$  does not have to be a correct lower bound of the cost of paths between  $n$  and  $t$ . An incorrect lower bound would imply larger values which could lead to excluding  $n$  from the search. However, since  $n \notin G_{SOL}(s, t)$ , it is not required to follow paths through  $n$ . We will start by defining the graph ParetoPrep visits when traversing  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$ :

**DEFINITION 6 (PARETOPREP GRAPH).** Let  $\mathcal{E}_{PP}(s, t) \subseteq \mathcal{E}$  denote the set of all edges which are considered in Step 4 of Algorithm 2. Analogously,  $\mathcal{V}_{PP}$  denotes the set of vertices of  $\mathcal{E}_{PP}(s, t)$  and  $\mathcal{C}_{PP}$  denotes the costs restricted to  $\mathcal{E}_{PP}(s, t)$ . Finally, we refer to  $G_{PP}(s, t) := (\mathcal{V}_{PP}, \mathcal{E}_{PP}(s, t), \mathcal{C}_{PP})$  as the ParetoPrep Graph.

Now, the first property is formulated and shown by the following lemma:

**LEMMA 4.** Given a path skyline query from  $s$  to  $t$  in a multicost network  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$ , it holds:  $G_{SOL} \subseteq G_{PP}$  (for the pair  $(s, t)$  which is omitted here for reasons of clarity).

**PROOF.** It suffices to show that  $\mathcal{E}_{SOL} \subseteq \mathcal{E}_{PP}$ . Then the claim follows by construction of the respective graphs. We show the proposed by contradiction: Suppose, there exists  $\mathcal{E}_{SOL} \ni (n_{i-1}, n_i) \notin \mathcal{E}_{PP}$ . Let  $p = ((s, n_1), \dots, (n_k, t)) \in \mathcal{S}$  denote a skyline path which contains edge  $(n_{i-1}, n_i)$ . Since  $(n_{i-1}, n_i) \notin \mathcal{E}_{PP}$ , the lower bound cost vector of some  $n_j, j \geq i$ , must have been globally dominated (cf. Step 3 in Algorithm 2). Since  $lb(n_j, t)_i \leq \text{cost}((n_j, n_{j+1}), \dots, (n_k, t))_i$  for all cost dimensions  $i$ , the subpath from  $n_j$  to  $t$  is also globally dominated. But the negation of Lemma 1 states: If a subpath is dominated, so is the whole path. Hence,  $p$  must be dominated which contradicts the assumption.  $\square$

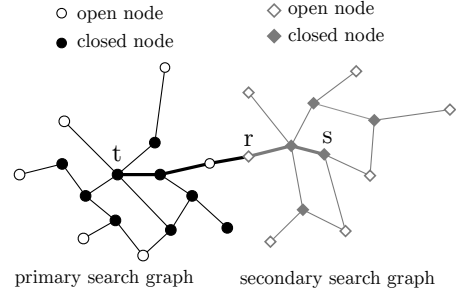
The second property we have to show is that the lower bounds for any node  $n \in G_{SOL}(s, t)$  are correct:

**LEMMA 5.** Given a path skyline query from  $s$  to  $t$  and its solution graph  $G_{SOL}(s, t)$ , let  $p_i(n, t)$  denote the shortest paths between node  $n \in G_{SOL}(s, t)$  and  $t$  for cost criterion  $i$  and let  $lb(n, t)_i$  be the lower bound value being computed by ParetoPrep for  $n$  and  $i$ . Then,  $lb(n, t)_i \leq \text{cost}(p_i(n, t))_i$ .

**PROOF.** Case (i):  $p_i(n, t) = ((n, n_1), \dots, (n_l, t))$  is the shortest path w.r.t. criterion  $i$  in both graphs  $G_{SOL}(s, t)$  and  $G_{PP}(s, t)$ . We now show that ParetoPrep traverses  $p_i(n, t)$  and thus,  $lb(n, t)_i = \text{cost}(p_i(n, t))_i$ . This is proven by induction.

**Basis:**  $(n_l, t)$  is traversed by ParetoPrep because any incoming node of  $t$  is examined. **Inductive step:** Given that  $((n_k, n_{k+1}), \dots, (n_l, t))$  is traversed, then  $lb(n_k, t)_i$  is updated and  $n_k$  is inserted into the open set. Because ParetoPrep does not terminate until the open set is empty,  $n_k$  will be processed at some iteration. Now, if  $lb(n_k, t)_i$  is currently globally nondominated in step 3, step 4 is performed and  $(n_k - 1, n_k)$  is examined which means the claim is proven. If  $lb(n_k, t)_i$  is currently globally dominated in step 3, we have to show that there has to be a later point in time where  $n_k$  is globally nondominated. Since  $n_k \in G_{SOL}(s, t)$  and  $G_{SOL}(s, t) \subseteq G_{PP}(s, t)$ , we know that  $n_k$  is guaranteed to be nondominated at the end of ParetoPrep. Thus, there must exist an iteration where  $n_k$  passes step 3 for the first time and step 4 is performed traversing the final edge  $(n_k - 1, n_k)$ . **Case (ii):**  $p_i(n, t)$  is the shortest path w.r.t. criterion  $i$  in  $G_{SOL}(s, t)$  but not w.r.t.  $G_{PP}(s, t)$ . In this case  $lb(n, t)_i < \text{cost}(p_i(n, t))_i$  holds because  $G_{SOL}(s, t) \subseteq G_{PP}(s, t)$  and the cost of any shortest path cannot be increased by extending the graph.  $\square$

After proving that ParetoPrep visits all relevant nodes  $n \in G_{SOL}(s, t)$  and computes valid lower bounds  $lb(n, t)$ , we show that all single criterion shortest paths between  $s$  and  $t$  for the complete graph are computed.



**Figure 6: Bidirectional ParetoPrep, when both its ParetoPrep searches rendezvous at node  $r$ . To reach  $s$  from outside the secondary search graph costs at least as much as the minimum of all  $lb^\beta$  cost vectors of the secondary search.**

**COROLLARY 6.** Given a multicost network  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$  and a path skyline query from  $s$  to  $t$ , it holds: For all cost dimensions  $i$   $lb(s, t)_i$  is the cost value of the shortest path from  $s$  to  $t$  (w.r.t. dimension  $i$ ) in the original graph  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$ .

**PROOF.** The single criterion shortest paths between  $s$  and  $t$  are part of  $\mathcal{S}(s, t)$  and also shortest paths in the complete graph  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$ . Thus, case (i) of the previous lemma can be applied.  $\square$

Hence, ParetoPrep computes all single criterion shortest paths between  $s$  and  $t$  (w.r.t.  $G(\mathcal{V}, \mathcal{E}, \mathcal{C})$ ). Also, ParetoPrep visits the subgraph that is relevant for answering a path skyline query and therein computes valid and tight lower bounds during one single traversal of  $G_{PP}$ .

### 4.3 Bidirectional ParetoPrep

Now that we have introduced the basic version of our algorithm and proven its theoretic foundation, we will introduce an optimized version of ParetoPrep. The idea of this optimization is the fact that the global domination check limiting the search space can still be improved. This is achieved by employing bidirectional search. In ParetoPrep we do not extend a node  $n$  if  $lb(n, t)$  is dominated by some path in  $\mathcal{S}$ . Though this check is correct, it is not tight in most cases because the cost of the path connecting  $s$  and  $n$  is disregarded. Thus, if we had a lower bound approximation for the costs of getting from  $s$  to  $n$   $lb(s, n)$ , the check for global domination in Step 3 of Algorithm 2 could be optimized by checking if  $\mathcal{S} \prec_{\text{dom}} lb(n, t) + lb(s, n)$ .

The idea of bidirectional ParetoPrep is to start two searches simultaneously. A backward search starts at  $t$  and proceeds in the same way as described above to derive lower bounds between the visited nodes and the target  $t$ . Additionally, a forward search starts at  $s$  and traverses the graph in the forward direction in order to find lower bounds for the path between  $s$  and intermediate nodes  $n$ .

At the beginning both searches take turns, i.e. iterations are alternated. Once both searches meet at some rendezvous node  $r$ , as shown in Figure 6, a path from  $s$  to  $t$  can be constructed for each cost dimension  $i$  by merging the current shortest paths from  $s$  to  $r$  and  $r$  to  $t$ . After that, only the backward search is continued.

In the following, we will refer to paths ending at nodes in the open set of the forward search as open paths. If  $t$  was not reached by a nondominated path  $p$  from  $s$  to  $t$ , there



has to be an open path, which is contained in  $p$ . If there is no such open path, ParetoPrep cannot reach  $s$  through  $p$ .

Let  $\text{mincost}^{\text{forward}}$  be the minimum cost of all open paths of the forward search. The cost vector  $\text{mincost}^{\text{forward}}$  consists of the lower bound costs of all globally nondominated paths from  $s$  to any node which has not been visited by the forward search. The minimum of all cost vectors of open nodes yields the minimum cost of all open paths:

$$\text{mincost}_i^\beta = \min_{o \in \text{open}_\beta} lb^\beta(o)_i$$

Let  $lb(s, n)$  denote the lower bound costs of all paths from  $s$  to  $n$  contained in a nondominated path from  $s$  to  $t$ . If no such costs are known, they have to be assumed as zero. If  $n$  was not reached by the forward search, they can be set to  $\text{mincost}^\beta$ :

$$lb(s, n)_i \geq \begin{cases} 0, & \text{if } n \text{ not visited by } \beta \text{ search} \\ \text{mincost}_i^\beta, & \text{if } n \text{ visited by } \beta \text{ search} \end{cases}$$

The cost vector  $\text{mincost}^{\text{forward}}$  is computed once when both searches rendezvous. The forward search is discontinued after the rendezvous. The cost vectors  $lb(s, n)$  are then used in the backward search to determine if a node was only reached by globally dominated paths. If that is the case, the node can be pruned.

In conclusion, this bidirectional ParetoPrep yields greater pruning power and thus, it restricts the search space even further.

## 5. EXPERIMENTS

**Experimental Setup:** The experiments we present in this section aim to show three things. Firstly, any label correcting search (LCS) for all skyline paths between designated start and target nodes in a network should utilize global domination. Secondly, our proposed algorithms ParetoPrep (PP) and Bidirectional ParetoPrep (BPP) outperform the state of the art approaches Doubledijkstra (DD) and Multidijkstra (MD) in terms of runtime as well as memory usage. Thirdly and finally, every LCS can be significantly accelerated in every way by employing the proposed preparatory search of ParetoPrep (PP) and even further accelerated by employing BPP. This also allows solving more complex tasks (e.g. multicost path skyline computation) in less time.

The LCS we implemented as comparative method is ARSC [12] (cf. Algorithm 1) but any other LCS yields similar results. The Reference Node Embedding, a preprocessing step which is presented in [12], is not used unless stated otherwise. Any lower bound costs are either set to zero or provided by a prior search like ParetoPrep. Note that any preprocessing step which is not executed at query time (like the Reference Node Embedding) cannot be applied to dynamic graphs. As mentioned before (cf. Section 3.3), for the special case of two cost criteria an optimized variant of ARSC is used which maintains skylines in sorted lists and utilizes the sort order when merging skylines [2, 22].

The experiments were conducted on the road network of the German state of Bavaria, provided by OpenStreet Maps and consisting of 1 023 561 nodes and 2 418 437 edges. There are two sets of routing tasks. The first set is comprised

of  $\binom{27}{2} \cdot 2 = 702$  routing tasks with Munich’s central station, Munich’s airport and Munich’s 25 city districts as end points. The second set is comprised of  $\binom{10}{2} \cdot 2 = 90$  routing tasks between the ten biggest cities in Bavaria.

In the experiments three variables are measured: the number of visited nodes, the number of the assembled paths and the runtime of the search task. The number of visited nodes by DD / MD / PP / BPP corresponds to the number of node entries and thereby reflects memory usage. The number of assembled paths by the LCS is also strongly correlated with memory usage since all paths not dominated by assembled paths have to be stored. The runtime is measured by letting all compared algorithms perform each task three times and taking the average. If a run takes longer than five minutes, it is aborted and counted as a time out. All experiments are conducted on a machine with an Intel Xeon E5-2609 (10MB Cache, 2.40 GHz, 1066 MHz FSB) and 32 GB RAM, running SUSE Linux 3.20.1 and Java OpenJDK IcedTea 1.7.0.09.

The cost criteria utilized in the experiments are travel duration ( $\text{dur}$ ), route length ( $\text{len}$ ), number of crossings ( $\text{cros}$ ), penalized travel duration ( $\text{dur}_p$ ) and energy loss ( $\text{ener}$ ). Criterion ( $\text{dur}$ ) assumes travel speeds to equal the speed limits whereas the penalized estimate ( $\text{dur}_p$ ) assumes additional 30 and 15 seconds for each crossing with and without traffic lights, respectively.

The energy loss estimate ( $\text{ener}$ ) is an synthetic model roughly derived from typical battery capacities of electric cars and their respective ranges. It incorporates altitude differences in the following sense: ascent increases the energy consumption by the gained potential energy and descent reduces the energy consumption (while negative values are corrected to zero). Let us stress that the authenticity of the cost models used has absolutely no influence on the computational benefits of our algorithms.

algorithm	millions of assembled paths	average runtime	time outs (>5min)
702 Munich tasks with $\text{dur}/\text{ener}$ (avg: 11.15 paths)			
LCS	0.29 (53.3×)	0.1s (54×)	0
LCS <sub>ss</sub>	15.95 (1.0×)	5.44s (1×)	0

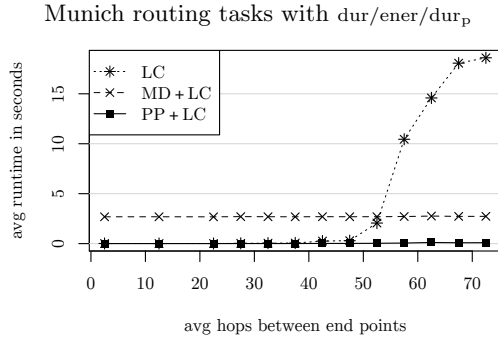
**Table 1: Comparison of point-to-point LCS and single-source LCS<sub>ss</sub>**

**The Impact of Global Domination:** Global domination pruning is indispensable for larger graphs and more difficult tasks. In this experiment only a part of the graph of Bavaria (around the city of Munich) with 221 465 nodes and 519 917 edges was used. This is because the single-source label correcting search (LCS<sub>ss</sub>) has to compute the nondominated paths from  $s$  to all other nodes in the graph which is not feasible for the complete graph of Bavaria. Even though restricting the graph assists the LCS<sub>ss</sub>, it can be observed that even in the smaller graph employing global domination checks reduces the number of assembled paths by a factor greater than 53 and leads to a reduction in runtime by the factor 54 (cf. Table 1). It should be noted that this is without utilizing lower bound costs which significantly improve the pruning power of global domination checks even further.

**Label Correcting Search and PP:** Utilizing lower bounds

algorithm	millions of assembled paths	average runtime	time outs (>5min)
90 Bavaria tasks with dur/ener (avg: 229.1 paths)			
<b>PP+LCS</b>	<b>5.2 (15.2×)</b>	<b>5.8s (6.9×)</b>	<b>0</b>
REF <sub>49</sub> +LCS	20.7 (3.8×)	14.4s (2.8×)	0
REF <sub>16</sub> +LCS	22.4 (3.5×)	13.5s (2.9×)	0
REF <sub>4</sub> +LCS	26.9 (2.9×)	14.9s (2.7×)	0
REF <sub>1</sub> +LCS	>48.1 (1.6×)	>30.3s (1.3×)	2
LCS	>79.6 (1.0×)	>40.5s (1.0×)	3
702 Munich tasks with dur/ener/dur <sub>p</sub> (avg: 125.9 paths)			
<b>PP+LCS</b>	<b>0.05 (38.2×)</b>	<b>1.4s (48.8×)</b>	<b>0</b>
MD+LCS	0.05 (38.2×)	4.0s (17.3×)	0
REF <sub>49</sub> +LCS	>0.44 (4.3×)	>14.4s (4.7×)	10.3
REF <sub>16</sub> +LCS	>0.46 (4.1×)	>16.3s (4.3×)	15
LCS	>1.93 (1.0×)	>70.4s (1.0×)	102

**Table 2: Comparison of runtime and number of assembled paths of LCS on its own and preceded by PP / MD or in combination with a reference point embedding with  $k$  evenly distributed landmarks (REF <sub>$k$</sub> +LCS).**



**Figure 7: Runtime comparison of LCS on its own and preceded by MD and PP for 290 local tasks with less than 70 hops between endpoints.**

in global domination checks significantly reduces the number of paths that have to be assembled. This improves runtime and memory usage of the search. The runtime of Pareto-Prep in these experiments is only 0.8s on average per query. This is only a small portion of the overall query time but the reduction of the label correcting runtime is about tenfold.

Preceding a LCS with PP and MD is crucial to solve more difficult routing tasks in a manageable time frame. As can be seen in Table 2 for 702 Munich tasks with dur/ener/dur<sub>p</sub>, a LCS on its own can solve less than one seventh of the Munich tasks in less than five minutes whereas with PP or MD all tasks are solved in less than two and a half minutes. On average LCS is at least 17 times slower than MD+LCS and at least 49 times slower than PP+LCS.

PP is clearly preferable to MD. As shown in Figure 7, MD+LCS is often slower than LCS on its own. PP+LCS, on the other hand, is always faster – on average 2.7 times faster than MD+LCS.

Table 2 also shows that a LCS in combination with PP is about twice as fast as with a Reference Node Embedding.

This result is very impressive since PP, unlike the Reference Node Embedding, does not require any precomputation and works instantly for any given graph. It is faster because the information provided by PP is superior to that provided by a Reference Node Embedding and thereby reduces the number of assembled paths during the LCS. PP also does not compute the minimum of lower bound vectors provided by different reference nodes, which is why 49 reference nodes lead to worse runtime results than 16 for two cost criteria although the overall number of assembled paths is smaller.

cost criteria	% of graph's nodes visited			
	LCS	DD/MD+LC	PP+LC	BPP+LC
90 Bavaria tasks				
dur/len	37.7 %	58 %	44.9 %	<b>30.8 %</b>
dur/ener	39.2 %	39.2 %	44.7 %	<b>32.8 %</b>
702 Munich tasks				
dur/ener/dur <sub>p</sub>	2.57 %	100 %	3.87 %	<b>2.1 %</b>

**Table 3: Comparison of visited nodes by an LCS on its own and when preceded by DD / MD, PP or BPP.**

Preceding a LCS by BPP reduces the number of visited nodes. As shown in Table 3, PP and DD visit more nodes than an LCS whereas BPP visits less.

	avg % of nodes visited (× less than DD)		
	DD	PP	BPP
702 Munich routing tasks			
dur/len	3.6 (1.0×)	2.7 (1.3×)	1.6 (2.3×)
dur/ener	3.6 (1.0×)	2.8 (1.3×)	1.8 (2.0×)
dur/dur <sub>p</sub>	5.4 (1.0×)	3.6 (1.5×)	1.7 (3.1×)
dur/cros	11 (1.0×)	4.5 (2.5×)	2.0 (5.6×)
90 Bavaria routing tasks			
dur/len	58 (1.0×)	44 (1.3×)	30 (1.9×)
dur/ener	56 (1.0×)	44 (1.3×)	32 (1.7×)
dur/dur <sub>p</sub>	45 (1.0×)	41 (1.1×)	26 (1.7×)
dur/cros	59 (1.0×)	46 (1.3×)	31 (1.9×)

**Table 4: Search area comparison of previous approach DD and proposed approaches PP and BPP.**

**Comparison of PP with DD:** This comparison is limited to two cost criteria, since DD cannot handle more than two cost criteria.

As can be seen in Table 4, PP and BPP visit less nodes than DD. For the Munich tasks BPP visits from two up to almost six times less nodes than DD and almost half as many nodes for the Bavaria tasks. Since there is one node entry for each visited node, BPP can be implemented noticeably more memory efficient than DD.

PP also outperforms DD in terms of runtime, as can be seen in Table 5. BPP is about twice as fast for the Munich tasks and about 1.5 times faster for the Bavaria tasks.

**Comparison of PP with MD:** Table 6 shows that PP and BPP visit only a small portion of the graph. This is ex-

	avg runtime in ms ( $\times$ faster than DD)		
	DD	PP	BPP
702 Munich routing tasks			
dur/len	58 (1.0 $\times$ )	43 (1.4 $\times$ )	32 (1.8 $\times$ )
dur/ener	53 (1.0 $\times$ )	43 (1.2 $\times$ )	36 (1.5 $\times$ )
dur/dur <sub>p</sub>	83 (1.0 $\times$ )	55 (1.5 $\times$ )	34 (2.4 $\times$ )
dur/cros	145 (1.0 $\times$ )	72 (2.0 $\times$ )	41 (3.5 $\times$ )
90 Bavaria routing tasks			
dur/len	1118 (1.0 $\times$ )	804 (1.4 $\times$ )	726 (1.5 $\times$ )
dur/ener	1119 (1.0 $\times$ )	883 (1.3 $\times$ )	816 (1.4 $\times$ )
dur/dur <sub>p</sub>	887 (1.0 $\times$ )	717 (1.2 $\times$ )	573 (1.5 $\times$ )
dur/cros	1057 (1.0 $\times$ )	945 (1.1 $\times$ )	768 (1.4 $\times$ )

**Table 5: Runtime comparison of previous approach DD and proposed approaches PP and BPP.**

	avg % of nodes visited ( $\times$ less than MD)		
	MD	PP	BPP
702 Munich routing tasks			
dur/ener/dur <sub>p</sub>	100 (1.0 $\times$ )	3.9 (25.8 $\times$ )	2.2 (46.0 $\times$ )
dur/len/cros	100 (1.0 $\times$ )	4.8 (20.7 $\times$ )	2.4 (41.9 $\times$ )
dur/ener/cros	100 (1.0 $\times$ )	4.8 (20.7 $\times$ )	2.5 (39.9 $\times$ )
all 5 criteria	100 (1.0 $\times$ )	4.7 (21.5 $\times$ )	2.4 (41.6 $\times$ )
90 Bavaria routing tasks			
dur/ener/dur <sub>p</sub>	100 (1.0 $\times$ )	47 (2.1 $\times$ )	35 (2.8 $\times$ )
dur/len/cros	100 (1.0 $\times$ )	52 (1.9 $\times$ )	38 (2.6 $\times$ )
dur/ener/cros	100 (1.0 $\times$ )	52 (1.9 $\times$ )	39 (2.5 $\times$ )
all 5 criteria	100 (1.0 $\times$ )	52 (1.9 $\times$ )	39 (2.5 $\times$ )

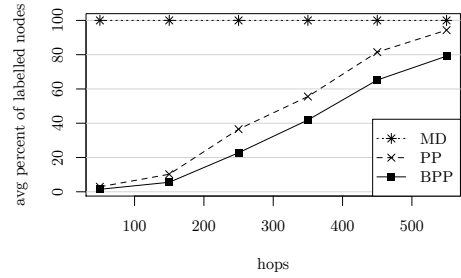
**Table 6: Search area comparison of previous approach MD and proposed approaches PP and BPP.**

tremely important if local searches in very large graphs have to be performed which is a common use case. Even when using five cost criteria, PP and BPP explore a fraction of the graph (cf. Figure 8). In contrast, MD always visits the complete graph once for each cost criterion. Figure 3 illustrates the visited area of the graph for one of these tasks and implies that BPP would also work well on gigantic graphs, e.g. consisting of all road networks in the world.

Table 7 shows that PP as well as BPP are much faster than MD. This of course is due to the limited search space. Using five criteria, BPP is 75.6 times faster than MD for the Munich tasks and 3.5 times faster for the Bavaria tasks. As can be seen in Figure 9 PP has the biggest runtime advantage for local search tasks.

## 6. CONCLUSION

In this paper, we introduce a new efficient algorithm to compute path skyline queries between two nodes in large multicriteria networks. Our methods works with multiple dimensions and are suitable for dynamic networks because they do not require any precomputed information. The core of the proposed is the algorithm ParetoPrep which efficiently computes all single criterion shortest paths between the two query nodes and additionally provides lower bound costs



**Figure 8: Percentage of visited nodes for different number of hops using Munich and Bavaria Routing tasks with all 5 criteria.**

	avg runtime in ms ( $\times$ faster than MD)		
	MD	PP	BPP
702 Munich routing tasks			
dur/ener/dur <sub>p</sub>	3082 (1.0 $\times$ )	88 (35.0 $\times$ )	58 (52.4 $\times$ )
dur/len/cros	3374 (1.0 $\times$ )	94 (35.5 $\times$ )	58 (57.7 $\times$ )
dur/ener/cros	3204 (1.0 $\times$ )	102 (31.2 $\times$ )	65 (49.3 $\times$ )
all 5 criteria	5323 (1.0 $\times$ )	112 (47.1 $\times$ )	70 (75.6 $\times$ )
90 Bavaria routing tasks			
dur/ener/dur <sub>p</sub>	3345 (1.0 $\times$ )	1272 (2.6 $\times$ )	1163 (2.9 $\times$ )
dur/len/cros	3299 (1.0 $\times$ )	1561 (2.1 $\times$ )	1301 (2.5 $\times$ )
dur/ener/cros	3344 (1.0 $\times$ )	1521 (2.2 $\times$ )	1382 (2.4 $\times$ )
all 5 criteria	5436 (1.0 $\times$ )	1690 (3.2 $\times$ )	1533 (3.5 $\times$ )

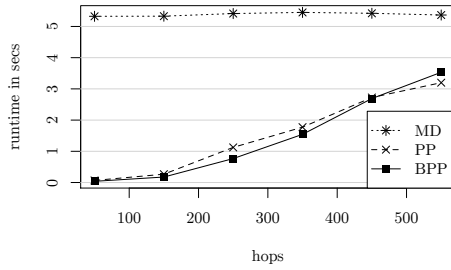
**Table 7: Runtime comparison of previous approach MD and proposed approaches PP and BPP.**

for the subgraph being relevant to the query. The information being computed by ParetoPrep can then be employed by a label correcting search algorithm like ARSC in two ways. First of all the set of single criterion shortest paths allows pruning by global domination from the start. Furthermore, the computed lower bounds provide tight forward estimations and thus, strongly help to reduce the number of constructed paths. We show that the results of ParetoPrep fulfill the formal requirements for this usage. Additionally, we introduce an optimized version of ParetoPrep, Bidirectional ParetoPrep that further limits the visited part of the graph and hence, increases the runtime efficiency of path skyline computation. In our experimental evaluation, we show that global domination is a key technique to allow fast path skyline computation. Furthermore, we demonstrate that ParetoPrep considerably outperforms other methods which compute lower bounds like Reference Node Embedding and the state of the art multiple Dijkstra searches.

For future work, we investigate more complex versions of multicost networks having uncertain and time-dependent cost values. Furthermore, we will investigate the combination of ParetoPrep with hierarchical networks for long distance skyline computation.

## 7. REFERENCES

- [1] S. Borzsonyi, D. Kossmann, and K. Stocker. The skyline operator. In *Proceedings of the 17th International Conference on Data Engineering (ICDE), Heidelberg, Germany, 2001*.



**Figure 9: Runtime for different number of hops using Munich and Bavaria Routing tasks with all 5 criteria.**

- [2] J. Brumbaugh-Smith and D. Shier. An empirical investigation of some bicriterion shortest path algorithms. *European Journal of Operational Research*, 43(2):216–224, 1989.
- [3] K. Deng, Y. Zhou, and H. T. Shen. Multi-source query processing in road networks. In *Proceedings of the 23th International Conference on Data Engineering (ICDE), Istanbul, Turkey*, 2007.
- [4] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.
- [5] M. Ehrgott and X. Gandibleux. A survey and annotated bibliography of multiobjective combinatorial optimization. *OR-Spektrum*, 22(4):425–460, 2000.
- [6] A. V. Goldberg and C. Harrelson. Computing the shortest path: A\* search meets graph theory. Technical Report MSR-TR-2004-24, Microsoft Research, 2004.
- [7] P. Hansen. Bicriterion path problems. In *Multiple criteria decision making theory and application*, pages 109–127. Springer, 1980.
- [8] X. Huang and C. S. Jensen. In-route skyline querying for location-based services. In *In Proc. of the Int. Workshop on Web and Wireless Geographical Information Systems (W2GIS), Goyang, Korea*, pages 120–135, 2004.
- [9] S. M. Jang and J. S. Yoo. Processing continuous skyline queries in road networks. In *International Symposium on Computer Science and its Applications (CSA2008)*, 2008.
- [10] D. Kossmann, F. Ramsak, and S. Rost. Shooting stars in the sky: an online algorithm for skyline queries. In *Proceedings of the 28th International Conference on Very Large Data Bases (VLDB), Hong Kong, China*, 2002.
- [11] H. P. Kriegel, M. Renz, and M. Schubert. Route skyline queries: a multi-preference path planning approach. In *Proceedings of the 26th International Conference on Data Engineering (ICDE), Long Beach, CA*, pages 261–272, 2010.
- [12] H.-P. Kriegel, M. Renz, and M. Schubert. Route skyline queries: A multi-preference path planning approach. In *Data Engineering (ICDE), 2010 IEEE 26th International Conference on*, pages 261–272. IEEE, 2010.
- [13] E. Machuca and L. Mandow. Multiobjective heuristic search in road maps. *Expert Syst. Appl.*, 39(7):6435–6445, June 2012.
- [14] E. Machuca and L. Mandow. Multiobjective heuristic search in road maps. *Expert Systems with Applications*, 39(7):6435–6445, 2012.
- [15] E. Q. V. Martins. On a multicriteria shortest path problem. *European Journal of Operational Research*, 16(2):236–245, 1984.
- [16] M. Müller-Hannemann and K. Weihe. Pareto shortest paths is often feasible in practice. In *Algorithm Engineering*, pages 185–197. Springer, 2001.
- [17] J. M. A. Pangilinan and G. JANSSENS. Evolutionary algorithms for the multi-objective shortest path problem. 2007.
- [18] D. Papadias, Y. Tao, G. Fu, and B. Seeger. An optimal and progressive algorithm for skyline queries. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), San Diego, CA*, 2003.
- [19] A. Raith and M. Ehrgott. A comparison of solution strategies for biobjective shortest path problems. *Computers & Operations Research*, 36(4):1299–1331, 2009.
- [20] M. Shekelyan, G. Jossé, M. Schubert, and K. H.-P. Linear path skyline computation in bicriteria networks. In *Proceedings of the 19th International Conference on Database Systems for Advanced Application (DASFAA), Bali, Indonesia*, pages 0–16, 2014.
- [21] A. J. Skriver. A classification of bicriterion shortest path (bsp) algorithms. *Asia Pacific Journal of Operational Research*, 17(2):199–212, 2000.
- [22] A. J. Skriver and K. A. Andersen. A label correcting approach for solving bicriterion shortest-path problems. *Computers & Operations Research*, 27(6):507–524, 2000.
- [23] B. S. Stewart and C. C. White III. Multiobjective a\*. *Journal of the ACM (JACM)*, 38(4):775–814, 1991.
- [24] K. L. Tan, P.-K. Eng, and B. C. Ooi. Efficient progressive skyline computation. In *Proceedings of the 27th International Conference on Very Large Data Bases (VLDB), Roma, Italy*, 2001.
- [25] Z. Tarapata. Selected multicriteria shortest path problems: An analysis of complexity, models and adaptation of standard algorithms. *Int. J. Appl. Math. Comput. Sci.*, 17(2):269–287, June 2007.
- [26] G. Tsaggouris and C. Zaroliagis. Multiobjective optimization: Improved fptas for shortest paths and non-linear objectives with applications. In *Algorithms and Computation*, pages 389–398. Springer, 2006.
- [27] C. Tung Tung and K. Lin Chew. A multicriteria pareto-optimal path algorithm. *European Journal of Operational Research*, 62(2):203–209, 1992.
- [28] A. Warburton. Approximation of pareto optima in multiple-objective, shortest-path problems. *Operations Research*, 35(1):70–79, 1987.
- [29] B. Yang, C. Guo, C. S. Jensen, M. Kaul, and S. Shang. Multi-cost optimal route planning under time-varying uncertainty. In *Proceedings of the 30th International Conference on Data Engineering (ICDE), Chicago, IL, USA*, 2014.